

37589



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In Application of : ZLOTNICK

Serial No.: 09/616,977

: Group Art Unit: 2178

Filed : July 14, 2000

: Examiner: Kyle R. Stork

For : DIRECTORY SERVICE FOR FORM PROCESSING

Honorable Commissioner for Patents

P.O. Box 1450

Alexandria, Virginia 22313-1450

DECLARATION UNDER 37 CFR 1.131

Sir:

I, the undersigned, Aviad Zlotnick, hereby declare as follows:

1) I am the Applicant in the patent application identified above, and am the sole inventor of the subject matter described and claimed in claims 1-37 therein.

2) Prior to March 24, 2000, I conceived my invention, as described and claimed in the subject application, in Israel, a WTO country. Conception of the invention is evidenced by an IBM Disclosure that I wrote, entitled "Internet Directory Service For Forms Processing" (serial no. 94850-1), which is attached hereto as Appendix A.

3) The dates deleted from Appendix A are prior to

US 09/616,977

Declaration under 37 C.F.R 1.131 by Zlotnick

March 24, 2000.

4) The following table shows the correspondence between the elements of claim 12 (as amended) in the present patent application and statements in the Disclosure attached as Appendix A:

Claim 12	Disclosure
A method for processing forms, each form including a field that is filled in with information in a predefined domain	Title: "Internet Directory Service For Forms Processing." Examples of domains include medical practice offices (page 1, second paragraph) and insurance (page 2, first unnumbered paragraph).
defining, in advance of reading out contents of the forms for processing, a directory of data relating to the predefined domain by selecting data specific to the domain from one or more general databases	"... an organization invests efforts in gathering directory information..." (page 1, third paragraph). "One could start with an established, purchased, database, and employ agents to find updates" (page 1, last paragraph). "It may be necessary for DS to go out and gather information in order to build all the directories needed by SI..." (page 2 first unnumbered paragraph).

US 09/616,977

Declaration under 37 C.F.R 1.131 by Zlotnick

Claim 12	Disclosure
receiving from a client via a computer network the information that is filled into the field on the forms by a plurality of users in communication with the client	"In particular, one could offer a service that accepts field images and context, and returns the field content in coded format (ASCII), or one could define an interface in coded format and return information in the same format" (page 2, first paragraph). "According to this contract, SI's system will send DS's web site field images, together with field classification" (page 2, first unnumbered paragraph).
checking whether the information is correct by looking up the information in the directory	"... extensive use of directory information can dramatically reduce the number of keystrokes needed for data entry from paper" (page 1, first paragraph). "... DS will respond by supplying verified OCR results" (page 2, first unnumbered paragraph). The verification is based on the directory information that the organization has gathered (page 1, first three paragraphs).

This table demonstrates that I conceived the entire

US 09/616,977

Declaration under 37 C.F.R 1.131 by Zlotnick

invention, as recited in claim 12, prior to March 24, 2000. Based on the similarity of subject matter between claims 12, 30 and 37, it can similarly be demonstrated that I conceived the entire invention recited in claims 30 and 37.

5) On March 14, 2000, I met with Dr. Daniel Kligler, of Sanford T. Colb & Co., who was retained by IBM as outside counsel for the purpose of preparing the present patent application. I was informed that Dr. Kligler had a substantial backlog of new applications that he was preparing for IBM, and that there would consequently be a delay of approximately two months in drafting this application. Immediately following the meeting, Dr. Kligler sent a memo to Tal Noy-Cohen, IP Manager of the IBM Haifa Research Laboratory, summarizing the meeting and timetable for completion of the draft application. A copy of this memo is attached hereto as Appendix B.

6) On May 8, 2000, I queried Dr. Kligler by e-mail as to the expected schedule for completion of a draft of this application, and he responded that the draft would be ready by the end of the month. A copy of the e-mail correspondence is attached hereto as Appendix C.

7) On May 30, 2000, Dr. Kligler sent me a first draft of the patent application. I responded immediately with comments and corrections to the draft. A copy of the draft with Dr. Kligler's cover letter is attached hereto as Appendix D.

8) On June 1, 2000, Dr. Kligler sent me a revised

US 09/616,977

Declaration under 37 C.F.R 1.131 by Zlotnick

draft of the patent application. A copy of Dr. Kligler's cover letter is attached hereto as Appendix E. The draft itself was identical to the patent application that was subsequently filed.

8) I immediately approved the revised draft for filing. It was then sent to an IBM in-house attorney, Jules Williams, for review. Mr. Williams gave final approval to file the application, and Ms. Noy-Cohen passed the approval on to Dr. Kligler by e-mail on June 20, 2000. A copy of this e-mail is attached hereto as Appendix F.

9) On June 25, 2000, Dr. Kligler's firm sent the patent application to Ms. Noy-Cohen together with documents for me to execute before filing. A copy of the cover letter under which the documents were sent is attached hereto as Appendix G.

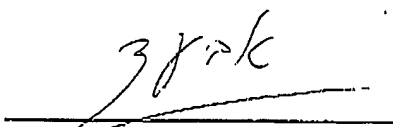
10) I executed the filing documents on July 2, 2000. A copy of the signature page of the Declaration is attached hereto as Appendix H. The application was then sent to the United States, where it was filed on July 14, 2000.

I hereby declare that all statements made herein of our my knowledge are true and that all statements made on information and conjecture are thought to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code

US 09/616,977

Declaration under 37 C.F.R 1.131 by Zlotnick

and that such willful false statements may jeopardize the validity of the application of any patent issued thereon.



Aviad Zlotnick, Citizen of Israel
Mizpe Netofa
D.N. Galil Takhton

Date:

Jun. 22, '06

IBM CONFIDENTIAL

37589

Inventor	Aviad Zlotnick
Serial	94850-1
Dept.	HRL
Zip	Haifa 31905
Location	Israel
Phone	972-4-990-8056

Internet Directory Service For Forms Processing

Background

In a recent test at a customer's site we showed that extensive use of directory information can dramatically reduce the number of keystrokes needed for data entry from paper. The reason for this is that many fields on typical forms relate to addresses, telephone numbers, and various identification codes. Using sophisticated directory lookup (fuzzy search) engines, it is possible to retrieve the content of all these fields even with OCR success on a small subset of the field characters.

Some directories, such as telephone directories, are readily available, at least in a version that is only almost up to date. Other directories are much harder to get. For instance, we wanted a directory of all the medical practice offices in the USA, and it was not available. In many forms processing applications such directories may change the economics of a solution.

This disclosure discusses a business model in which an organization invests efforts in gathering directory information, and makes profits by selling services related to this information via the internet. This model fits in well with IBM's recent policy of emphasizing technology and services.

It should be noted that the some of these same services are useful even when data capture is done directly through internet forms. One still could benefit from eliminating typos, and shortening the data gathering sessions.

Patent protection is sought for the business model, content free framework, and software for this business.

The Business Model

The business model has four components:

1. Information gathering: each organization can use its own ideas for information gathering. One could start with an established, purchased, database, and employ agents to find updates. Or one could build a database using fully internal resources. In most cases it

- would be desirable to maintain to database up to date, using whatever means possible.
2. Interface: the interface defines what information the customers provide, what they get back, and how. In particular, one could offer a service that accepts field images and context, and returns the field content in coded format (ASCII), or one could define an interface in coded format and return information in the same format. I think there is use for both services.
 3. Directory lookup: the search engines used may make a big difference in the quality of service, whether it is the OCR engine or the fuzzy search algorithm. A complete service may even include manual verification or manual key in.
 4. Payment method: payment for the services can be done by transaction - pay per field, or by project - a fixed price for the duration of the project. Here too, there seems to be use for both types of payment.

As an example, let us think of a directory service provider DS, and a software integrator SI. SI wants to automate data collection for an insurance company, but does not have expertise in OCR. Instead of developing OCR technologies from scratch, or purchasing off the shelf packages and starting to learn their particulars, SI goes to DS and signs a service contract. According to this contract SI's system will send DS's web site field images, together with field classification, and DS will respond by supplying verified OCR results. It may be necessary for DS to go out and gather information in order to build all the directories needed by SI, but with some luck, after doing business with several software integrators, DS will have most of the databases ready.

As mentioned above, DS may decide to code all of SI's transactions manually. As long as the response time, throughput and price are acceptable for SI, the business will run smoothly.

State of the Art

No such services exist in the document processing market.

In the internet domain, search sites like Yahoo and Alta Vista provide the same kind of service (information compilation and sophisticated search), but on a word basis instead of on a character basis. The internet search model is also different in that one cannot negotiate for special databases, and the payment model is different.

Advantages

Every organization does what it knows best. System integration people do not have to get into image or text processing, and computer science experts do not have to deal with user interfaces and hardware. The interface overhead should not be forbidding in this kind of computing intensive service.

In particular, this type of service makes it possible to build document processing systems in places where the volumes are too low to justify the investment in a standard system.

IBM CONFIDENTIAL

Date: March 16, 2000

To: Adv. Tal Noy-Cohen, IBM

From: Daniel Kligler
Sanford T. Colb & Co.

Re: IL9-2000-0009 - our 37589 - estimate of time and charges

Title: Internet directory service for document processing

Inventors: Aviad Zlotnick

Meeting held: March 14, 2000

Materials received: Invention disclosure

Time est.: First draft by mid-May

Cost est.: \$4,000 + VAT in professional fees, not including filing costs or out-of-pocket expenses.

Comments: A patent search turned up no relevant prior art.

F A X E D

Daniel Kligler

From: Daniel Kligler [dkligler@stc.co.il]
Sent: Tuesday, May 09, 2000 7:46 AM
To: 'aviad@il.ibm.com'
Subject: RE: Internet directory service - our ref. 37589

Dear Aviad,

A number of other IBM applications pushed in line ahead of this one, including your own application on image expansion and decimation. I still have one or two other IBM applications ahead of it in line, but I expect to send you a draft by the end of this month.

Regards and hag sameah,
Danny

-----Original Message-----

From: aviad@il.ibm.com [SMTP:aviad@il.ibm.com]
Sent: Mon, May 08, 2000 5:03 PM
To: dkligler@stc.co.il
Subject:

Danny,

It's a long time since I heard about my invention "An Internet Directory Service for Document Processing". I'd like to make sure it has not been lost. The PDT was held on March 14.

Aviad



Sanford T. Colb & Co.
Intellectual Property Law

Beit Amot Mishpat
8 Shaul Hamelech Blvd.
Tel-Aviv 64733, Israel
Tel. 972-3-693-8560

4 Shaar Hagai
P.O. Box 2273
Rehovot 76122, Israel
Tel. 972-8-945-5122

Beit Lev Hagivah
11 Beit Hadfus
Jerusalem 95483, Israel
Tel. 972-2-651-9453

Beit Topaz, 3rd Floor
Shaar Hacarmel
Haifa, Israel
Tel. 972-4-8503444

Facsimile: 972-8-945-4556 972-8-949-1040 ♦ e-mail: colbpat@stc.co.il

IBM CONFIDENTIAL

May 30, 2000

Mr. Aviad Zlotnick
IBM ISRAEL
Haifa Research Laboratory
MATAM, Haifa 31905

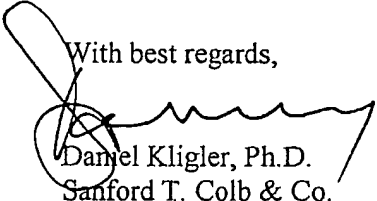
Re: New U.S. patent application
DIRECTORY SERVICE FOR FORM PROCESSING
Your ref. IL9-2000-0009, our ref. 37589

Dear Aviad:

Attached please find a first draft of the above-referenced patent application.

Please review this draft, and let us have your corrections and comments at your earliest opportunity. Note particularly a question that I have marked in **boldface** in the text of the application.

With best regards,


Daniel Kligler, Ph.D.
Sanford T. Colb & Co.

encl.

cc: Adv. Tal Noy-Cohen

DIRECTORY SERVICE FOR FORM PROCESSING**FIELD OF THE INVENTION**

The present invention relates generally to computerized information processing, and specifically to
5 extracting data from filled-in form documents.

BACKGROUND OF THE INVENTION

Methods for extraction of information filled into form documents are well known in the art. Typically, a document is printed with a form template. The template
10 contains predefined fields that are filled in by a user with appropriate characters. The document is scanned into a computer, which typically uses an optical character recognition (OCR) program to identify and code the characters in each field.

15 OCR identification of handwritten, or even typed, characters can be uncertain, due to a range of problems including uneven scan quality, variable character shapes, and interference between the filled-in characters and features of the printed template. A variety of methods
20 and systems have been developed to deal with these problems. For example, U.S. Patents 5,182,656, 5,191,525 and 5,793,887, whose disclosures are incorporated herein by reference, describe methods for registering a document image with a form template so as to remove the template
25 and extract the filled-in information from the form. Once the form is accurately registered with the known template, it is a simple matter for the computer to assign the fill-in characters to the appropriate fields. Dropping the template from the document image also

reduces substantially the volume of memory required to transmit or store the image.

Because of the uncertainty of machine identification of characters by OCR, methods have been developed for selectively verifying the correctness of coded results. For example, U.S. Patent 5,455,875, whose disclosure is incorporated herein by reference, describes a system and method for correction of optical character recognition, based on an interactive display of OCR results that is designed to enable an operator to correct erroneous character data reliably and efficiently.

Even in data that are not generated by OCR, there are commonly errors and inconsistencies, such as address information that is out of date or misspelled. To deal with problems of this sort, a number of companies offer address verification services, in which a mailing list is checked against an up-to-date master list. One example of such a service is "InfoBase BestAddress," offered by Acxiom Corporation, as described at www.acxiom.com. This service both identifies incorrect addresses and, where possible, provides corrections. The U.S. Postal Service offers master address databases that can be used to do this sort of verification.

SUMMARY OF THE INVENTION

In preferred embodiments of the present invention, a directory service receives information extracted from a form that has been filled in by a user. The information
5 is typically sent to the directory service via a computer network by a client, who has received the filled-in form from the user and needs the information contained in one or more fields on the form to be coded and verified. The service returns the coded and verified results to the
10 client over the network. Typically, multiple fields on multiple copies of the form, filled in by different users, are processed in this manner.

To deal with the information that is to be sent by the client, the directory service defines and assembles a
15 directory of data that is specific to a domain or category to which the information belongs. Preferably, the service assembles the specific directory by culling the data from other, more general databases. The service codes the information filled into the form, and then
20 looks up the coded information in the directory to verify that the information is coded correctly and/or to choose among a number of possible codes when the coding is uncertain. The use of the specific, focused directory enables the service to search and verify the coded
25 information with greater reliability and speed than are generally achievable with general-purpose databases, such as public-domain telephone and address listings.

In some preferred embodiments of the present invention, the users fill in the forms by writing or
30 typing characters into the fields. Preferably, the client sends images of the filled-in field to the service

via the network, and the service uses OCR techniques to code the characters. Alternatively, the client may itself code the characters in the field and then send the coded results, or a number of alternative codes, to the service. In either case, by verifying the OCR output against the directory, the service is able to identify and eliminate errors in the OCR coding and to reduce the number of uncertain OCR readings that need to be passed to a human operator for verification. Thus, by using the directory service, a client who is not expert in OCR and does not have convenient access to appropriate, focused directories is able to obtain high-quality coding results without a major investment in acquiring new infrastructure or capabilities.

Preferably, the client pays the service for providing the coded information on the basis of the quantity of information that is processed. Most preferably, the payment is calculated based upon a price per field processed. Alternatively, the payment may be on the basis of processing resources, such as CPU time, expended in coding and verifying the information, or on a fixed price or subscription basis, or on substantially any other commercial basis that is known in the art.

{Claim summary will be inserted here.}

The present invention will be more fully understood from the following detailed description of the preferred embodiments thereof, taken together with the drawings in which:

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram that schematically illustrates a system for processing information filled into forms, in accordance with a preferred embodiment of the present invention;

Fig. 2 is a flow chart that schematically illustrates a method for building a directory, in accordance with a preferred embodiment of the present invention; and

Fig. 3 is a flow chart that schematically illustrates a method for processing information filled into a form, in accordance with a preferred embodiment of the present invention.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

Fig. 1 is a block diagram that schematically illustrates a system 20 for processing information filled into a form 24, in accordance with a preferred embodiment of the present invention. In the scenario shown in Fig. 1, a client 22, such as a system integrator, is responsible for automating data collection from a large number of forms, but does not have in house the capabilities needed to process the data automatically. Rather than purchasing software and developing the necessary capabilities, which would require a large investment of time and capital, client 22 contracts with a directory service 30 to perform the processing. The directory service typically comprises one or more suitable computer processors with software for carrying out the methods described hereinbelow. The software may be furnished to the directory service in electronic form, via a network or other link, or it may be supplied on tangible media, such as CD-ROM or non-volatile memory.

Each filled-in form received by client 22 is scanned by a scanner 26 to form an electronic image of the form, as is known in the art. The client sends the entire form image or selected elements of the image, as described hereinbelow, to service 30 via a computer network 28, typically via the Internet. The directory service applies OCR to code the characters filled into the form, and then uses one or more directories 32 stored in a memory or other storage device 33 to verify that the coding is correct. For example, assuming form 24 to be a medical insurance form, which includes fields for the name and address of a treating physician, the directory

service would preferably procure or produce a directory of physicians against which to verify this information. After completing the coding and verification process, service 30 returns the coded results via network 28 to client 22.

Fig. 2 is a flow chart that schematically illustrates a method by which directory service 30 assembles the directory needed for a particular verification job, in accordance with a preferred embodiment of the present invention. Together with client 22, the directory service defines a domain over which the information in form 24 is to be searched, at a search definition step 34. This domain might be the population of practicing physicians in the United States, for example.

At the same time, the directory service receives a definition of the specific fields that are to be coded, at a field definition step 36. In the case of the insurance form mentioned above, for example, these fields might include the physician's name, address and specialization, as well as an identification of the patient and the procedure carried out. The client and directory service preferably also agree at this stage as to the form in which the field contents for processing are to be sent from the client to the service. Preferably, the client sends electronic images of the fields, which are to be coded by the service using OCR. Alternatively, the field contents may be sent to the service already in coded form. This will be the case, for example, when the client itself performs the OCR (thereby reducing the volume of data that must be sent

over network 28) or when the forms have been filled in electronically, so that OCR is not required. Although in this latter case the directory service no longer needs to deal with OCR coding errors, directory lookup is still
5 useful in detecting and correcting typographical errors and other inaccuracies.

Based on the domain and field definitions, the directory service preferably assembles a special-purpose directory for use in verifying the results of coding the
10 filled-in forms, at a directory building step 38. Preferably, the directory service purchases and maintains a stock of specialized databases, such as the physician directory mentioned above. Alternatively or
15 additionally, the directory service builds and maintains directories of its own, typically by assembling information from general, public-domain databases and from other available sources. Further alternatively or
20 additionally, general databases, such as postal or telephone directories, may be used when appropriate. Most preferably, the directory service employs agents and surveys sources of information to keep its directories up to date.

Fig. 3 is a flow chart that schematically illustrates a method for processing the information in
25 form 24 by directory service 30, in accordance with a preferred embodiment of the present invention. This method uses the field definitions and directory generated at steps 36 and 38, as described above. The description of the method of Fig. 3 assumes that client 22 receives
30 paper forms, comprising a template filled in by users with handwritten or printed characters. The method is

also applicable, however, *mutatis mutandis*, to forms that are filled in electronically.

Each form 24 that is received by client 22 is scanned to generate an electronic image of the form, at a form input step 40. Preferably, a template registration and drop-out program, as is known in the art, is provided on the client's computer in order to register the image with a template of the form and to remove the template from the image. Suitable methods for this purpose are described, for example, in the above-mentioned U.S. Patents 5,182,656, 5,191,525 and 5,793,887. Removal of the template reduces the volume of information that must be transmitted over network 28 to directory service 30 and makes subsequent OCR processing easier and more accurate. Alternatively, client 22 transmits the entire image to service 30, and template drop-out is performed by the service or not at all.

Following template drop-out, the fields to be coded by the directory service are located on the form, at a field identification step 44. The identification is typically based on predefined positions of the fields in the form template. Preferably, this step, as well, is performed by suitable software operated by client 22, whereby only the images of the specific fields of interest are transmitted subsequently to service 30. Alternatively, the appropriate fields for processing are extracted from the overall image by the directory service.

The images of the selected fields are read and coded, at a content reading step 46. Any suitable method of OCR that is known in the art may be used at this step

(assuming that form 24 is a paper form, whose content must be coded). Preferably, the OCR program returns one or more possible readings of the content, each with a respective confidence score. The results of the coding are verified against the data in the selected directory, at a lookup step 48. When step 46 returned only a single reading, step 48 is used to confirm that the coded contents agree with one of the entries in the directory (for example, that the physician's name, address and specialty all match). Preferably, a "fuzzy," error-tolerant search algorithm is used, so that small errors, such as misspellings or OCR misreadings, can be detected and overcome, without leading to rejection of an otherwise valid coding result. {Can you give an example, preferably a patent, that describes the fuzzy search that you would use here?} When multiple, alternate readings are suggested by step 46, the directory lookup at step 48 is used to choose the most likely reading among the alternatives.

Step 48 thus either confirms or modifies the coding result generated at step 46. Preferably, the confidence score from step 46 is also modified by step 48, typically increasing the confidence level to "certain" when an OCR reading is found to correspond with high likelihood to an entry in the directory. On the other hand, when the OCR reading does not correspond to any directory entry, its confidence level may be reduced. At a confidence checking step 50, the confidence level of the coding result is compared to a predetermined threshold. If the confidence is below threshold, the original field is passed to a human operator, preferably together with the

(uncertain) coding results, at a manual coding step 52. Any suitable method of data presentation may be used to assist the operator in processing the information efficiently and reliably, such as that described in U.S. Patent 5,455,875. The operator either confirms or selects the appropriate coding result from among the alternatives offered by the OCR, or enters a different, correct result.

The verified coding result for each field is returned to client 22 at a concluding step 54. Preferably, the directory service charges the client for its work on the basis of the number of fields, words or characters that have been processed. Alternatively, the charge may be based on a fixed, periodic payment, or on a measure of use of the resources of the directory service, such as CPU time, or on substantially any other payment basis known in the art.

While preferred embodiments described herein relate particularly to form documents and OCR coding, it will be understood that the principles of the present invention are similarly applicable to verification of data coding generated by other methods and to processing documents of other types. It will thus be appreciated that the preferred embodiments described above are cited by way of example, and that the present invention is not limited to what has been particularly shown and described hereinabove. Rather, the scope of the present invention includes both combinations and subcombinations of the various features described hereinabove, as well as variations and modifications thereof which would occur to

37589S

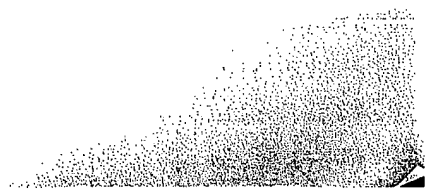
IBM CONFIDENTIAL

persons skilled in the art upon reading the foregoing description and which are not disclosed in the prior art.

IL9-2000-0009

12

37589s 05/30/00 9:06 AM DK



CLAIMS

1. A method for processing a document including a field containing information in a predefined domain, the method comprising:
 - 5 defining a directory of data relating to the predefined domain;
receiving from a client via a computer network an image of the field containing the information;
processing the image to code the information; and
 - 10 looking up the coded information in the directory so as to verify that the information is coded correctly.
2. A method according to claim 1, and comprising returning the verified, coded information over the network to the client.
- 15 3. A method according to claim 2, wherein receiving the image of the field comprises receiving a number of fields filled in with respective information, regarding which the verified, coded information is returned to the client, and comprising receiving payment from the client
- 20 according to the number of the fields.
4. A method according to claim 1, wherein defining the directory comprises selecting data specific to the predefined domain from one or more general databases.
5. A method according to claim 1, wherein receiving the
- 25 image comprises receiving an image of alphanumeric characters in the field.
6. A method according to claim 5, wherein the document includes a template delineating the field, and wherein receiving the image of the characters comprises receiving

the image of the characters filled into the field and remaining after drop-out of the template from the image of the field.

7. A method according to claim 5, wherein processing
5 the image comprises applying computerized optical character recognition (OCR) to code the characters.

8. A method according to claim 7, wherein looking up
the coded information comprises selecting a preferred
reading of the characters from among two or more possible
10 readings generated by the OCR, responsive to the data in the directory.

9. A method according to claim 7, wherein looking up
the coded information comprises generating a confidence
score, and wherein processing the image comprises passing
15 the image to a human operator for coding when the confidence score is below a predetermined threshold.

10. A method for processing forms, each form including a
field that is filled in with information in a predefined
domain, the method comprising:

20 defining a directory of data relating to the predefined domain by selecting data specific to the domain from one or more general databases;

receiving from a client via a computer network the
information that is filled into the field on the forms by
25 a plurality of users in communication with the client;
and

verifying correctness of the information by looking
up the information in the directory.

11. A method according to claim 10, wherein receiving
30 the information comprises receiving coded information,

and wherein verifying the correctness comprises verifying that the coded information is correct.

12. A method according to claim 11, wherein receiving the coded information comprises receiving coded
5 characters generated by the client using optical character recognition (OCR).

13. A method according to claim 10, wherein receiving the information comprises receiving an image of the field, and comprising processing the image to code the
10 information, wherein verifying the correctness of the information comprises verifying that the information was coded correctly by looking up the coded information in the directory.

14. A method according to claim 10, and comprising
15 returning the verified information over the network to the client.

15. A method according to claim 14, and comprising receiving payment from the client according to a number of the forms for which the correctness of the information
20 in the field was verified.

16. Apparatus for processing a document including a field containing information in a predefined domain, the apparatus comprising:

25 a memory, in which a directory of data relating to the predefined domain is stored; and

a directory service processor, adapted to receive from a client via a computer network an image of the field containing the information, to process the image to code the information, and to look up the coded

information in the directory so as to verify that the information is coded correctly.

17. Apparatus according to claim 16, wherein the processor is adapted to return the verified, coded
5 information over the network to the client.

18. Apparatus according to claim 17, wherein the processor is adapted to receive a number of fields filled in with respective information, regarding which it is to return verified, coded information, and to receive
10 payment from the client according to the number of the fields.

19. Apparatus according to claim 16, wherein the directory comprises data specific to the predefined domain, which are selected from one or more general
15 databases.

20. Apparatus according to claim 16, wherein the image comprises alphanumeric characters filled into the field.

21. Apparatus according to claim 20, wherein the document includes a template delineating the field, and
20 wherein the characters in the image comprise the characters remaining after drop-out of the template from the image of the field.

22. Apparatus according to claim 20, wherein the processor is adapted to apply computerized optical
25 character recognition (OCR) to code the characters.

23. Apparatus according to claim 22, wherein the processor is further adapted to select a preferred reading of the characters from among two or more possible

readings generated by the OCR, responsive to the data in the directory.

24. Apparatus according to claim 22, wherein the processor is further adapted to generate a confidence score in a reading generated by the OCR, and to pass the image to a human operator for coding when the confidence score is below a predetermined threshold.

25. Apparatus for processing forms, each form including a field that is filled in with information in a predefined domain, the apparatus comprising:

a memory, in which a directory of data relating to the predefined domain is stored by selecting data specific to the domain from one or more general databases; and

a processor, adapted to receive from a client via a computer network the information that is filled into the field on the forms by a plurality of users in communication with the client, and to verify correctness of the information by looking up the information in the directory.

26. Apparatus according to claim 25, wherein the processor is adapted to receive coded information, and to verify that the coded information is correct.

27. Apparatus according to claim 25, wherein the processor is adapted to receive an image of the field and to process the image to code the information, wherein the processor is adapted to verify that the information was coded correctly by looking up the coded information in the directory.

28. Apparatus according to claim 25, wherein the processor is adapted to return the verified information over the network to the client.

29. A computer software product for processing a document including a field that contains information in a predefined domain, the product comprising a computer-readable medium in which program instructions are stored, which instructions, when read by a computer, cause the computer to receive a definition of a directory of data relating to the predefined domain and, upon receiving from a client via a computer network an image of the field containing the information, to process the image so as to code the information and to look up the coded information in the directory so as to verify that the information is coded correctly.

30. A product according to claim 29, wherein the image comprises alphanumeric characters filled into the field, and wherein the instructions cause the computer to apply optical character recognition (OCR) to code the characters.

31. A computer software product for processing forms, each form including a field that is filled in with information in a predefined domain, the product comprising a computer-readable medium in which program instructions are stored, which instructions, when read by a computer, cause the computer to receive a definition of a directory of data relating to the predefined domain generated by selecting data specific to the domain from one or more general databases, and upon receiving from a client via a computer network the information that is

37589S

IBM CONFIDENTIAL

filled into the field on the forms by a plurality of users in communication with the client, to verify correctness of the information by looking up the information in the directory.



Sanford T. Colb & Co.
Intellectual Property Law

F A X E D

Beit Amot Mishpat
8 Shaul Hamelech Blvd.
Tel-Aviv 64733, Israel
Tel. 972-3-693-8560

4 Shaar Hagai
P.O. Box 2273
Rehovot 76122, Israel
Tel. 972-8-945-5122

Beit Lev Hagivah
11 Beit Hadfus
Jerusalem 95483, Israel
Tel. 972-2-651-9453

Facsimile: 972-8-945-4556 972-8-949-1040 ♦ e-mail: colbpat@stc.co.il

IBM CONFIDENTIAL
38 pages via fax

June 1, 2000

Mr. Aviad Zlotnick
IBM ISRAEL
Haifa Research Laboratory
MATAM, Haifa 31905

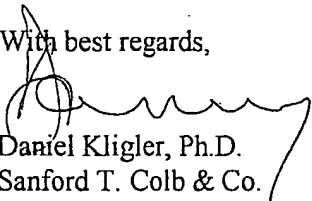
Re: New U.S. patent application
 DIRECTORY SERVICE FOR FORM PROCESSING
 Your ref. IL9-2000-0009, our ref. 37589

Dear Aviad:

Attached please find a revised draft of the above-referenced patent application, incorporating your comments and corrections. As you will see, I have changed occurrences of the term "verification" to "checking," and I have added claims on the error correction function. I hope this answers the reservation that you raised in your e-mail to me. In preparation for filing, I have also added an abstract and paraphrased the claims in the Summary of the Invention. For your convenience in reviewing this draft, I am attaching some of the pages with the changes marked.

Please review this draft, and let us have your approval to file the application at your earliest opportunity. Please pass a copy to Tal Noy-Cohen, as well, for her to send on to Jules Williams for review before filing.

With best regards,


Daniel Kligler, Ph.D.
Sanford T. Colb & Co.

encl.

cc: Adv. Tal Noy-Cohen

Daniel Kligler

APPENDIX F

From: noy@il.ibm.com
Sent: Tuesday, June 20, 2000 11:41 AM
To: dkligler@stc.co.il
Cc: hrl_ipdepartment@il.ibm.com
Subject: 37589, IL9-2000-0009 (IS899-0088)

Daniel, Please file. Thanks
Tal

Tal Noy - Cohen, adv.
Voice: +972-4-829-6274
Fax: +972-4-829-6521
Intellectual Property Coordinator
Intellectual Property Department
Operations Services, IBM HRL

משרד רחובות: ת.ד. 2273, רחובות 76122, מל. 08-9455122, פקס. 08-9454556, 08-9491040
משרד חלאיב: בית אמות משפס, שד' שאול הסנך 8, חל-אביב 64733, מל. 03-6938560, פקס. 6938561
משרד ירושלים: לב הנכבה, רחוב בית הרפס וו, ירושלים 95483, מל. 02-6519453, פקס. 02-6519454
משרד חיפה: בית טופס, קומה שלישיח, שער הכרמל, חיפה, מל. 04-8503444, פקס. 04-8503555
REHOVOT OFFICE: P.O.B. 2273, REHOVOT, TEL. 08-9455122, TELECOPIER 08-9454556, 08-9491040
TEL-AVIV OFFICE: BEIT AMOT MISHPAT, 8 SHAUL HAMELECH BOULEVARD
64733 TEL-AVIV, ISRAEL, TEL. 03-6938560, TELECOPIER 03-6938561
JERUSALEM OFFICE: LEV HAGIVAH, 11 BEIT HADFUS, JERUSALEM, TEL. 02-6519453, TELECOPIER 02-6519454
HAIFA OFFICE: BEIT TOPAZ, 3RD FLOOR, SHAAH HACARMEI, HAIFA, TEL. 04-8503444, TELECOPIER 04-8503555
E-Mail: colbpaf@stc.co.il

BY SPECIAL MAIL DELIVERY

Re: New U.S. Patent Application
 DIRECTORY SERVICE FOR FORM PROCESSING
 Your Ref.: IL9-2000-0009
 Our Ref.: 37589

Jean Levy
Foreign Filing Department
Sanford T. Colb & Co.

9/7/52

APPENDIX H

FULL NAME AND RESIDENCE OF INVENTOR 2

LAST NAME: FIRST NAME: MIDDLE NAME:
CITY: STATE OR FOREIGN COUNTRY:
COUNTRY OF CITIZENSHIP:
POST OFFICE ADDRESS: CITY: STATE OR COUNTRY:
ZIP CODE:

FULL NAME AND RESIDENCE OF INVENTOR 3

LAST NAME: FIRST NAME: MIDDLE NAME:
CITY: STATE OR FOREIGN COUNTRY:
COUNTRY OF CITIZENSHIP:
POST OFFICE ADDRESS: CITY: STATE OR COUNTRY:
ZIP CODE:

I further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

SIGNATURE OF INVENTOR 1: X 3/r/c DATED: X July 2, '00

SIGNATURE OF INVENTOR 2: _____ DATED: _____

SIGNATURE OF INVENTOR 3: _____ DATED: _____